

## Predictive Modelling of Breast and Oral Cancer Incidence Trends in Sri Lanka: Utilizing Random Forest Algorithm and National Cancer Registry Data

D Satharasinghe<sup>1#</sup>, A Neththikumara<sup>1</sup>, J Jeyasugiththan<sup>1</sup>, L Liyanage<sup>2</sup>, RMDS Rasnayake<sup>3</sup>, and Oshada Senaweera<sup>4</sup>

<sup>1</sup>Department of Nuclear Science, Faculty of Science, University of Colombo, Sri Lanka

<sup>2</sup>Department of Statistics and Computer Science, Faculty of Science, Humanities and Social Sciences, University of Kelaniya, Sri Lanka

<sup>3</sup>Department of Engineering Technology, Faculty of Technology, General Sir John Kotelwala Defence University, Sri Lanka

<sup>4</sup>Department of Statistics, Faculty of Science, University of Colombo, Sri Lanka

#duminda@nuclear.cmb.ac.lk

Predicting cancer incidence trends is essential for effective healthcare planning and resource allocation. This study utilized the Random Forest model to forecast cancer patterns in Sri Lanka using comprehensive data from the National Cancer Registry spanning 2001 to 2020. The model's performance was rigorously evaluated, demonstrating robust predictive capabilities with accuracies of 89.0% for female breast cancer and 88.0% for male oral cancers. Projections extend to 2030, revealing an upward trajectory in cancer cases, with notable annual variations, particularly in breast cancer among females post-2024. Despite challenges such as limited data availability from private sector sources and variability in diagnostic practices across regions, the study underscores the transformative potential of machine learning in enhancing cancer surveillance and management strategies. By integrating medical and socio-demographic factors into predictive frameworks, this research contributes to advancing understanding of cancer epidemiology in developing regions like Sri Lanka. Moving forward, efforts should focus on refining models through enhanced data integration and addressing modifiable risk factors such as lifestyle choices and healthcare accessibility disparities. This study emphasizes the critical role of predictive modelling in shaping proactive public health policies and interventions aimed at mitigating the increasing cancer burden in Sri Lanka and in similar global contexts.

**Keywords:** *cancer prediction, random forest, cancer incidence, oral cancer, breast cancer*